



HAL
open science

Group Emotion Recognition Strategies for Entertainment Robots

Sarah Cosentino, Estelle Randria, Jia-Yeu Lin, Thomas Pellegrini, Salvatore
Sessa, Atsuo Takanishi

► **To cite this version:**

Sarah Cosentino, Estelle Randria, Jia-Yeu Lin, Thomas Pellegrini, Salvatore Sessa, et al.. Group Emotion Recognition Strategies for Entertainment Robots. IEEE/RSJ International Conference on Intelligent RObots and Systems - IROS 2018, Oct 2018, Madrid, Spain. pp.813-818, 10.1109/IROS.2018.8593503 . hal-02982961

HAL Id: hal-02982961

<https://hal.science/hal-02982961>

Submitted on 6 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Group emotion recognition strategies for entertainment robots

Sarah Cosentino^{1§}, Estelle I. S. Randria^{2§}, Jia-Yeu Lin¹, Thomas Pellegrini³,
Salvatore Sessa¹, Atsuo Takanishi^{1,4,5}, *Member, IEEE*

Abstract— In this paper, a system to determine the emotion of a group of people via facial expression analysis is proposed for the Waseda Entertainment Robots.

General models and standard methods for emotion definition and recognition are briefly described, as well as strategies for computing the group global emotion, knowing the individual emotions of group members. This work is based on Ekman's extended "Big Six" emotional model, popular in Computer Science and Affective Computing. Emotion recognition via facial expression analysis is performed with a cloud-computing based solution, using Microsoft Azure Cognitive services. First, the performances of both the Face API to detect faces, and Emotion API, to compute emotion via face expression analysis, are tested. After that, a solution to compute the emotion of a group of people has been implemented and its performances compared to human perceptions. This work presents concepts and strategies which can be generalized for applications within the scope of assistive robotics and, more broadly, affective computing, wherever it will be necessary to determine the emotion of a group of people.

Keywords—humanoid robot; entertainment robot; assistive robotics; emotion recognition

I. INTRODUCTION

Human social communication is complex, and comprehends two communication channels: rational and emotional. The first one is conscious and result of cognitive, rational processes. The second one is often unconscious and the result of emotional unconscious processes [1].

Social communication is also adaptive, its evolution depending on both the rational and the emotional responses of the counterpart [2]. In particular, rational communication uses a predefined, structured language, verbal or non-verbal, while emotional communication largely relies on body language, vocal paralinguistics, and facial expressions [2]–[4]. Both channels are equally important: if rational exchanges provide an idea of the intentions and expectations of both interacting parties, emotional exchanges provide context and social distance information [5].

In previous research, it has been found that people also treat and respond to modern media devices in the same way as they treat and respond to other people in everyday social

interaction [6]. In fact, modern media devices are treated as social agents, and the rules which people apply to everyday social interaction with other people apply equally well to their interactions with these devices. For this reason, as the next generations of robots are expected to navigate and perform tasks in human-centered environments, participating and supporting humans in daily routines, they are expected to be social robots capable of interacting with humans in a natural way. Thus, social assistive robots should learn not only how to communicate rationally, using a predefined, structured language, but also emotionally, recognizing and responding to subtle and sometimes ambiguous emotional signals.

Most importantly, socially assistive robots will often have to deal with more than one user at once, so they should be able to identify the general emotional context in which multiple interactions are performed, and, if needed, to assign a priority level to each single interaction and to first choose and respond to higher priority ones. However, even if emotional analysis has been a long-studied problem in affective computing for its importance in human-robot interaction, research focused on inferring the emotional state of a single subject only. Analyzing the general emotion of a group of people is an unexplored problem that has many implications, especially for assistive robots [7].

The objective of this research is to verify the feasibility of a system able to recognize the emotions of a group of people and compute the group general emotion, to enable an assistive robot to generate in response tailored emotional feedback.

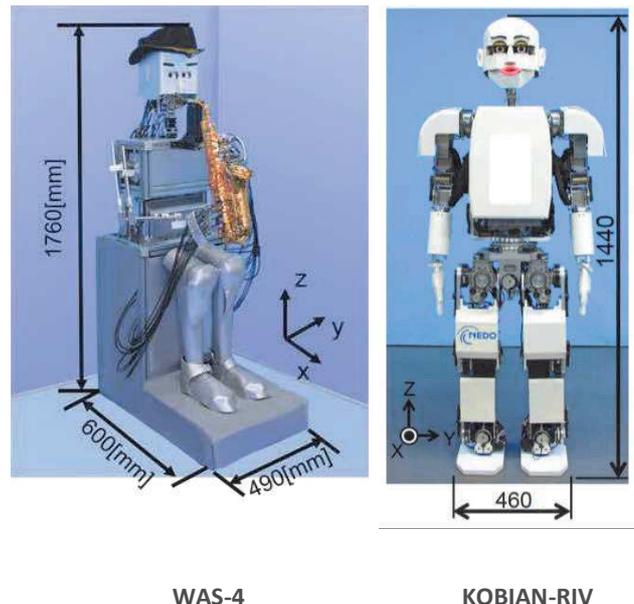


Figure 1 Waseda Humanoid Entertainment Robots

*Research supported by the Waseda University 2016 Grant-in-Aid for particular research subjects [2016S-081] and by the JSPS Grant-in-Aid for Young Scientists (Wakate B) [17K18178]; by ST Microelectronics, and Microsoft.

§Both authors contributed equally to this manuscript

1 Faculty of Science and Engineering, Waseda University, Tokyo, Japan

2 Université Paul Sabatier (Upssitech), Toulouse, France.

3 Université de Toulouse, UPS, IRIT, Toulouse, France.

4 Humanoid Robotics Institute (HRI), Waseda University, Japan

5 Member IEEE (email: contact@takanishi.mech.waseda.ac.jp)

More specifically, in this work:

- 1) a system to track and analyze simultaneously the emotions of many people has been developed and tested
- 2) automatic strategies to evaluate the general emotion of a group of people, provided the individual emotions of most people in the group are estimated correctly, have been implemented and compared with human strategies.

This paper is organized as follows: in Section II the tools and methods used for the detection and interpretation of emotions are presented, together with the issue of establishing criteria to model the emotion of a group of people; experiments and results are presented in Section III, followed by a discussion in Section IV, and conclusions and future research directions in the same field in Section V.

II. MATERIALS AND METHODS

A. Waseda social robot platforms

Entertainment robotics is a subgroup of socially assistive robotics. Entertainment robots are not made for utilitarian use, but for the pleasure and well-being of the user. They can be useful in a wide range of applications, from the educational to the healthcare field [8], [9]. Entertainment robots must be able to recognize, adapt to and possibly positively influence the emotional state of the user(s). The Waseda Anthropomorphic Saxophonist (WAS) and KOBIAN humanoid robots are both an example of entertainment robotics (Figure 1).

WAS-4 is a musical humanoid robot, its dimensions are 30% larger than an average human [10]. KOBIAN-RIV is a purely social humanoid, its dimensions are comparable to an average Japanese female [11]. It is currently used as a comedian robot to explore the effects of amusement on the user long-term emotional state [12]. Both robots must be able to execute and interpret gaze, body movements, and facial expressions to communicate emotionally with their audience.

Most importantly, they should be able to analyze simultaneously the emotions of a group of people, compute the group emotion, and learn how to respond appropriately, with the objective of generating pleasure and increase the well-being of their audience. To be able to do so, the robots must rely on a framework to define and identify emotions and develop psychological strategies for appropriate emotional feedback.

In this section, the definition of emotions on which this research is based, as well as the chosen emotion recognition method are presented, explained, and validated.

B. Emotion definition

As Scherer writes, there is still no generally accepted definition of emotion [13]. However, it is accepted by now that emotions are states of interrelated, synchronized physiologic and behavioral changes in response to an external or internal stimulus [14]–[16]. Emotions are short-term mental states provoked by a particular stimulus, whilst a pervasive and sustained emotional state is called mood [16]. Psychologists have extensively studied emotions and emotional states, and several different models have been devised. These models try to define which emotions can be considered basic, as innate, unique, distinct

psychophysiological states, and which ones can instead be considered a mix of two or more basic emotions. Determining basic emotions is a complex task, as deciding the criteria for defining an emotion as basic are difficult to settle [17], [18]. To cite the most important models:

- Plutchik considered eight basic emotions: acceptance, anger, anticipation, disgust, joy, fear, sadness, surprise. He elaborated the well-known Plutchik’s wheel of emotions [19];
- Ekman first considered six basic emotions, the “Big Six”: happiness, sadness, fear, surprise, anger, and disgust. He then expanded his model to seven basic emotions, adding contempt, and eventually to fifteen, with amusement, embarrassment, excitement, guilt, pride in achievement, relief, satisfaction, sensory pleasure and shame [20].
- Weiner and Graham considered as basic emotions only sadness and happiness [21].
- James considered as basic emotions fear, grief, love and rage [22].

Psychological frameworks also conceptualize emotion along two dimensions, *valence* and *arousal*. Valence describes the extent of pleasure or displeasure, and arousal describes the extent of physiological excitement [19], [23]–[26]. Of the basic seven, including contempt, happiness only has a positive valence (pleasurable state), surprise has none (neutral), while all the other have negative valence (Table I).

C. Emotions recognition system

As emotions are linked to specific psychophysiological states, there are different methods to monitor and measure physiological parameters to estimate the linked emotion. However, the most ecological methods rely on audio and visual information. As WAS-4 and KOBIAN-RIV and in general most of the existing assistive robots are equipped with front cameras to detect and identify users, the proposed system relies on emotion recognition from visual cues. In particular, facial expressions are the most direct and clear visual display of emotions [27]. Following this approach, Ekman and Friesen developed the Facial Action Coding System (FACS) [15], a system to taxonomize human facial movements, and linked each emotion in Ekman’s model to a specific set of facial movements. Most of the successful works on automatic emotion recognition are based on facial expressions analysis and on Ekman’s basic or extended emotions models, and so is the proposed system.

Moreover, as the system will be used for recognition of the global emotion of the robot audience, it is expected to analyze simultaneously facial features of several people, and this will involve a high computational load. To solve this problem and easily scale computational power and expressions database, cloud computing has been used. In particular, the system is based on a cloud computing solution using Microsoft Azure, as this cloud computing service already contains two APIs for face detection and emotion detection: Microsoft Cognitive Services – *Face API* and *Emotion API* [28]. *Face API* analyzes an image and returns the number and the position of the faces in the image. *Emotion API* analyzes a face image and returns

TABLE I. VALENCE, EMOTIONS, AND AN EXAMPLE OF ARBITRARILY DEFINED EMOTION HIERARCHY (1: HIGH)

| Valence | Emotion | Priority |
|----------|-----------|----------|
| Negative | Sadness | 1 |
| Negative | Anger | 2 |
| Negative | Fear | 3 |
| Negative | Disgust | 4 |
| Negative | Contempt | 5 |
| Positive | Happiness | 6 |
| None | Surprise | 7 |
| None | Neutral | 8 |

coefficients in the range [0-1] for all the seven basic emotions of Ekman's model, plus neutral (no emotion). The higher the coefficient value, the stronger is the corresponding emotional component in the facial expression. The displayed emotion can be estimated using these coefficients, and in the simplest case, the estimated emotion equals the emotion with the highest coefficient.

D. Strategies for defining group emotion and mood

Defining the emotion or the mood of an audience composed by many people is not a simple task and several strategies can be used.

For example, the general emotion of the audience can be defined as the main emotion perceived: the emotion of each single person in the audience can be estimated, and the emotion with the highest occurrence selected. This strategy can lead to ambiguity problems when dealing with large audience where different emotions are present in equal number.

Another possibility, instead, is to define an emotion hierarchy. By arbitrarily prioritizing some emotions over others, when one person displays a higher priority emotion, this is chosen as the audience general emotion. This approach is convenient when addressing specific emotional states, and also to automatically resolve possible ambiguity problems (Table I).

A third possibility is to take into consideration not the emotions per se, but the overall emotional valence, the mood, of the estimated emotional states. This is a simpler strategy that reduces the problem to a three-state choice. As happiness is the only emotion with positive valence in Ekman's basic model, this is also a useful simplified strategy if the ultimate objective is to identify happiness, or a positive mood.

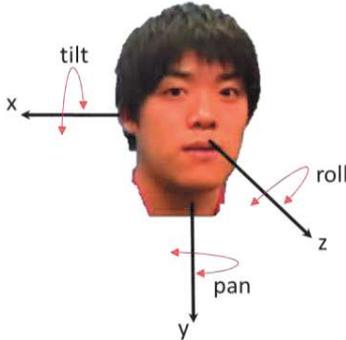


Figure 2 Tilt, Roll and Pan angles

An important point is to check how these strategies compare with human perception of a group emotion. It is reasonable to expect that, to appear as natural as possible, a social assistive robot should respond to a group in the same way in which a human would do in the same context. However, human strategies to determine group emotion, given individual emotions in the group are identified, are unknown, and little to no literature is available. To have an idea on how to model, and possibly reproduce, human strategies to determine group emotion, human perception of group emotion must be tested and analyzed.

III. EXPERIMENTS AND RESULTS

A. Face detection

The first step to examine the feasibility of our system is to test the performances of the system in tracking and analyzing simultaneously the emotions of many people. The performances of Face API and Emotion API have been tested separately. The performances in face detection of Face API are strongly influenced by the orientation of the face compared to camera. Orientation of an object in front of the camera is expressed by tilt (angle around the horizontal axis) roll (angle around the frontal axis) and pan (angle around the vertical axis) as shown in Figure 2. As the audience might move during the robot performance, pan might vary accordingly, and being the cameras fixed on the robot, tilt will result when the height of camera and the person face are mismatched. To test performances depending on tilt and pan variations an existing facial recognition database has been used, the Head Pose Image database [29]. This database is a benchmark of 2790 face images of 15 persons with variations of pan and tilt angles from -90 to +90 degrees. Specifically, variations are in the ranges:

- Tilt = {-90, -60, -30, -15, 0, +15, +30, +60, +90}
- Pan = {-90, -75, -60, -45, -30, -15, 0, +15, +30, +45, +60, +75, +90}

Results showed that face detection is over 90% within the range [-30°, +30°] of both tilt and pan (Figure 3).

It can be observed that tilt is more critical for recognition than pan, so the camera should have adjustable height for a more reliable recognition. In fact, if there is no tilt, face detection rate is a steady 100% in the pan range [-45°, +45°].

The API has also been tested with a sample of 50 pictures of frontal faces covered with various accessories, manually excerpted from the web. It has been found that accessories that cover part of the face like hat, sunglasses, or scarfs, strongly influence face detection, with extremely variable results and a recognition rate under 30%. Accessories that cover the lower part of the face are critical and do not allow the face detection with this API.

B. Emotion recognition

The performances of the Emotion API have been tested with the Karolinska Directed Emotional Faces (KDEF) database [30]. This database consists of 4900 pictures of human facial expressions of emotion with variations of pan angles. The database contains two sets of images of 70 actors, displaying the 6 basic emotional expressions, plus neutral –contempt is excluded– photographed from 5 different angles:

- Pan = {-90, -45, 0, +45, +90}
- Emotions = {anger,disgust,fear,happiness,neutral,sadness,surprise}

Results show that pan angles have also an influence on recognition rate of emotions, decreasing the correct recognition rate of a few percentage points (Figure 4).

In Figure 5 the emotional recognition confusion matrix for all frontal images from KDEF database is shown. It is clear that the Emotion API is finely tuned to recognized neutral, surprise and happiness expressions, thus neutral and positive emotional valence, but weak in differentiating basic emotions with a negative valence, especially fear, which is either

| PAN TILT | -90 | -75 | -60 | -45 | -30 | -15 | 0 | 15 | 30 | 45 | 60 | 75 | 90 |
|-------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|----|----|
| 90 | x | x | x | x | x | x | 3 | x | x | x | x | x | x |
| 60 | 3 | 3 | 13 | 53 | 53 | 80 | 83 | 73 | 67 | 33 | 10 | 3 | 0 |
| 30 | 10 | 43 | 77 | 93 | 100 | 100 | 100 | 100 | 93 | 83 | 70 | 37 | 17 |
| 15 | 23 | 53 | 90 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 77 | 53 | 30 |
| 0 | 33 | 60 | 93 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 70 | 50 | 17 |
| -15 | 17 | 40 | 87 | 100 | 97 | 100 | 100 | 100 | 100 | 90 | 60 | 23 | 7 |
| -30 | 7 | 20 | 83 | 93 | 97 | 93 | 100 | 97 | 93 | 80 | 40 | 17 | 3 |
| -60 | 0 | 7 | 27 | 37 | 47 | 63 | 73 | 77 | 53 | 47 | 17 | 0 | 3 |
| -90 | x | x | x | x | x | x | 20 | x | x | x | x | x | x |

Figure 3 Face recognition rate for tilt and pan combinations [%]

| EMOTION \ PAN | -90 | -45 | 0 | 45 | 90 |
|---------------|-----|-----|-----|-----|----|
| Anger | x | 60 | 82 | 60 | x |
| Disgust | x | 24 | 29 | 22 | x |
| Fear | x | 27 | 42 | 29 | x |
| Happiness | x | 100 | 100 | 100 | x |
| Neutral | x | 100 | 100 | 100 | x |
| Sadness | x | 96 | 97 | 96 | x |
| Surprise | x | 91 | 94 | 83 | x |

Figure 4 Emotion recognition rate depending on pan angle [%]

| KDEF \ RECOGN | Anger | Contempt | Disgust | Fear | Happiness | Neutral | Sadness | Surprise |
|---------------|-------|-------------------------|---------|------|-----------|---------|---------|----------|
| Anger | 55 | | 7.8 | 1.4 | | | | |
| Contempt | 2.8 | Not represented in KDEF | | 2.8 | | | | |
| Disgust | 2.1 | | 71.4 | 0.7 | | | | |
| Fear | 0.7 | | | 17.1 | | | | |
| Happiness | | | 2.1 | 4.2 | 100 | | 0.7 | 0.7 |
| Neutral | 35.7 | | 4.2 | 10.7 | | 100 | 12.1 | 3.6 |
| Sadness | 2.8 | | 14.3 | 20.7 | | | 86.4 | |
| Surprise | 0.7 | | | 42.8 | | | 0.7 | 95 |

Figure 5 Emotion recognition confusion matrix [%]

misclassified as surprise, sadness, or even neutral. This should be corrected, as fear is not an uncommon emotion evoked by human-sized humanoid robots, and robots must be able to recognize and appropriately respond to it.

C. Group emotion evaluation

To simulate multiple emotion recognition from facial expression situation, 6 frontal pictures extracted randomly from the KDEF have been combined (Figure 6). 25 images containing 6 different emotion faces have then been created with this method and used as sample to test the performances of the automatic emotion recognition system against human perception.

An online survey has been created and distributed to reach as many people as possible and to ensure diversity in the subjects' pool. In total, 25 people of age in the range of 20-40 y.o., 45.5% female-to-male ratio, from several countries (mostly Asian and European), and of different education backgrounds and levels, participated in the survey. Results of the survey can be seen in Figure 7, compared with the results of automatic group emotion estimation performed by the system on the same images. As could be expected, very few questions led to a unanimous 100% response on a single emotion.

Using Emotion API there are two methods of computing the global group emotion:

- 1) for each face, the main emotion can be computed from the highest emotional coefficient, so that 6 emotions can be determined first, then the group emotion is equaled to the emotion with the highest occurrence in the group
- 2) the relative emotional coefficients for all faces can be summed up for each emotion, obtaining a set of global emotional coefficients for the group, and the group image main emotion can be defined as the emotion with highest global coefficient

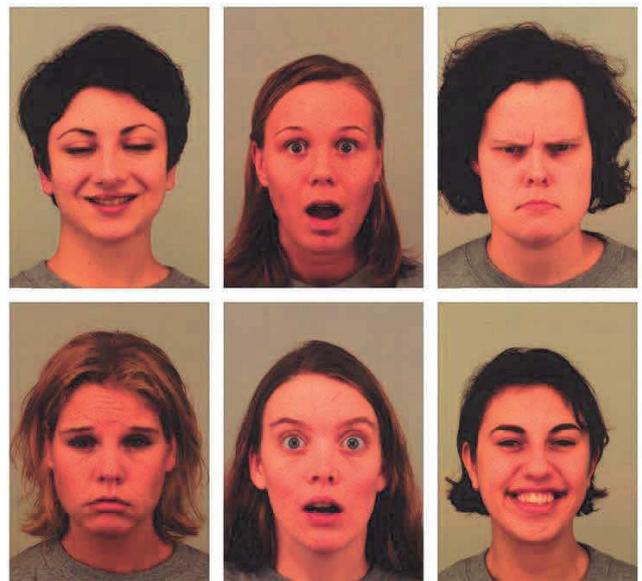


Figure 6 KDEF six emotional faces image sample (survey question 1)

Method 1) can easily lead to ambiguity problems when different emotions are present in equal occurrence, whilst method 2) is more reliable and can provide also a set of global emotional coefficients to use for more advanced recognition techniques, so this is the computational method used in the proposed system. Results are shown in Figure 7: human responses for each group image are shown in percentages, whilst for automatic recognition using Emotion API, for each group image the computed global coefficients are shown. To be noted, as contempt is not represented in KDEF, survey answers did not include contempt, but Emotion API computes the contempt coefficients anyway, obtaining as expected very low values. In the table, the computed contempt coefficients are in light grey. Highest responses in both situations are shaded in dark grey, second higher values in light grey.

| QUESTION# | RESPONSE | Human [%] | | | | | | | Automatic, global coefficients [%, approximated] | | | | | | | |
|-----------|----------|-----------|----|----|-----|-----|----|----|--|----|----|----|----|----|----|----|
| | | AN | DI | FE | HA | NE | SA | SU | AN | CO | DI | FE | HA | NE | SA | SU |
| 1 | | 0 | 4 | 0 | 4 | 8 | 4 | 80 | 16 | 0 | 0 | 0 | 34 | 3 | 14 | 33 |
| 2 | | 0 | 12 | 4 | 4 | 40 | 40 | 0 | 0 | 0 | 0 | 0 | 33 | 39 | 27 | 1 |
| 3 | | 24 | 44 | 4 | 16 | 8 | 4 | 0 | 17 | 1 | 14 | 0 | 34 | 33 | 2 | 0 |
| 4 | | 0 | 20 | 28 | 0 | 32 | 16 | 4 | 23 | 1 | 3 | 6 | 2 | 44 | 18 | 3 |
| 5 | | 0 | 0 | 0 | 32 | 4 | 0 | 64 | 0 | 0 | 0 | 0 | 52 | 1 | 0 | 47 |
| 6 | | 0 | 48 | 12 | 12 | 0 | 4 | 24 | 1 | 0 | 5 | 19 | 33 | 14 | 8 | 19 |
| 7 | | 0 | 8 | 4 | 0 | 56 | 32 | 0 | 1 | 0 | 0 | 2 | 17 | 41 | 24 | 15 |
| 8 | | 0 | 0 | 0 | 92 | 4 | 4 | 0 | 0 | 1 | 0 | 0 | 67 | 5 | 27 | 0 |
| 9 | | 0 | 0 | 0 | 0 | 20 | 8 | 72 | 0 | 1 | 0 | 1 | 1 | 40 | 14 | 43 |
| 10 | | 12 | 48 | 0 | 0 | 0 | 4 | 36 | 13 | 1 | 14 | 0 | 0 | 1 | 21 | 50 |
| 11 | | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 5 | 9 | 1 | 3 | 67 | 3 | 0 | 12 |
| 12 | | 0 | 16 | 4 | 0 | 4 | 76 | 0 | 0 | 0 | 0 | 0 | 33 | 2 | 65 | 0 |
| 13 | | 0 | 0 | 0 | 4 | 92 | 4 | 0 | 0 | 0 | 0 | 0 | 17 | 67 | 17 | 0 |
| 14 | | 0 | 0 | 4 | 96 | 0 | 0 | 0 | 0 | 0 | 1 | 6 | 50 | 32 | 8 | 2 |
| 15 | | 56 | 8 | 4 | 0 | 0 | 0 | 32 | 15 | 1 | 3 | 1 | 0 | 28 | 19 | 33 |
| 16 | | 20 | 8 | 20 | 0 | 0 | 0 | 52 | 13 | 0 | 0 | 18 | 0 | 32 | 8 | 28 |
| 17 | | 0 | 24 | 12 | 0 | 8 | 52 | 4 | 1 | 1 | 3 | 1 | 1 | 42 | 27 | 24 |
| 18 | | 88 | 8 | 0 | 0 | 0 | 0 | 4 | 32 | 13 | 13 | 0 | 17 | 14 | 9 | 1 |
| 19 | | 0 | 0 | 0 | 4 | 76 | 0 | 20 | 0 | 0 | 0 | 0 | 17 | 50 | 0 | 32 |
| 20 | | 0 | 12 | 4 | 0 | 0 | 84 | 0 | 1 | 2 | 1 | 0 | 23 | 12 | 61 | 0 |
| 21 | | 0 | 4 | 0 | 72 | 12 | 0 | 12 | 0 | 8 | 4 | 7 | 50 | 19 | 5 | 7 |
| 22 | | 4 | 40 | 0 | 0 | 0 | 0 | 56 | 6 | 1 | 42 | 0 | 0 | 1 | 1 | 49 |
| 23 | | 0 | 0 | 0 | 72 | 0 | 0 | 28 | 0 | 0 | 0 | 1 | 50 | 0 | 0 | 49 |
| 24 | | 0 | 36 | 20 | 40 | 0 | 4 | 0 | 0 | 0 | 2 | 15 | 60 | 1 | 18 | 3 |
| 25 | | 0 | 0 | 0 | 0 | 100 | 0 | 0 | 0 | 0 | 0 | 0 | 17 | 83 | 0 | 0 |

AN = Anger, CO = Contempt, DI = Disgust, FE = Fear, HA = Happiness, NE = Neutral, SA = Sadness, SU = Surprise

Figure 7 six emotional faces perception

Matching responses in automatic and human perception are circled. The automatic system recognition matches the majority of human responses for each group on 17 cases out of 25, with a majority matching rate of 68%.

IV. DISCUSSION

Observing the results, several considerations can be done.

In the real-time multiple faces detection stage, the main limitation is to track faces with tilt and pan in a certain range. Detection errors could be reduced placing the camera in a position allowing filming the most people possible regardless of their height and position with respect to the robot. For a tall robot this could be implemented placing multiple cameras at different heights, having the robot move around often to check faces in its surroundings, or with a movable camera. To have information about tilt and pan during recognition would also provide useful, as this could be passed on to the second stage of emotional recognition and provide information on possible emotion recognition errors. Emotional coefficient could be then weighted differently to adjust for tilt and pan induced errors.

The performances of the emotional recognition stage should be improved, especially toward the recognition rate of emotions with negative valence, anger, disgust and fear. The automatic system failed to recognize disgust and anger in nearly all the cases, except for group 18, composed by 1 happy, 3 angry, and 2 disgusted expressions. Coefficient normalization could be implemented to reduce the bias of the Emotion API towards happiness and neutral.

An important point of this research is that, as it is shown in Figure 7, a simple computational approach to determine the global emotion of a group cannot be used. In the presence of different emotional cues from different people, some subjects are more sensitive to positive emotions; some subjects are more sensitive to negative emotions. Moreover, depending on the actual situation context, it is probably appropriate to establish an emotional hierarchy, as theorized in Section II, to respond quickly to some contextually more important emotional cues. For example, in the case of the robot in a public setting, it could be appropriate to respond to anger or fear cues by receding and rolling away, to preserve the robot integrity or to avoid distress to the interacting human subject. Also, if the social assistive robot is expected to react to a group in a similar way a human would, in the same context, the robot should be able to approximate the human global group emotion perception, with a certain degree of uncertainty, and to be able to respond appropriately to mixed emotional states. Emotional responses should be tailored on both the group mood, positive or negative, and individual emotional cues, to avoid degrading the group mood and at the same time improving individual emotional states.

V. CONCLUSION

This work is a preliminary study on recognition of group emotions for assistive social robots. Basic theoretical concepts and design choices have been explained and analyzed, and an automatic computational approach has been compared with human perception. In the future, more human perception data should be acquired to train a machine learning system, to

develop a group emotion assessment paradigm similar to the one humans use, and teach the robot to respond appropriately to mixed group emotions, enhancing the decision power of the robot and its ability to entertain and engage its users. However, before implementing a magic black box emotional response tool, it is important to analyze in deep the mechanisms of human group emotional interaction, and to identify and contextualize the key strategies humans use during emotional interactions, to devise a method for qualitative and quantitative assessment of the automatic recognition system and identify and correct its possible biases.

ACKNOWLEDGMENT

The authors would like to thank the Tokyo Women Medical University and Waseda University Joint Institution for Advanced Biomedical Sciences (TWIns) and the International Center for Science and Engineering Program (ICSEP) of Waseda University.

REFERENCES

- [1] D. Morris and G. Desebrock, *Manwatching: A field guide to human behaviour*. HN Abrams New York, 1977.
- [2] J. M. Darley and R. H. Fazio, "Expectancy confirmation processes arising in the social interaction sequence," *Am. Psychol.*, vol. 35, no. 10, pp. 867–881, 1980.
- [3] M. J. Snyder, "Interpersonal processes: The interplay of cognitive, motivational, and behavioral activities in..." *Annu. Rev. Psychol.*, vol. 50, no. 1, p. 273, Feb. 1999.
- [4] R. R. Provine, *Curious Behavior: Yawning, Laughing, Hiccupping, and Beyond*. Harvard University Press, 2012.
- [5] N. Schwarz, "Emotion, cognition, and decision making," *Cogn. Emot.*, vol. 14, no. 4, pp. 433–440, Jul. 2000.
- [6] C. Nass and B. Reeves, *The media equation: How people treat computers, televisions, and new media as real people and places*. Cambridge University Press, 1996.
- [7] A. Dhall, J. Joshi, I. Radwan, and R. Goecke, "Finding Happiest Moments in a Social Context," in *Computer Vision – ACCV 2012*, vol. 7725, K. M. Lee, Y. Matsushita, J. M. Rehg, and Z. Hu, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 613–626.
- [8] J. Wainer, D. J. Feil-Seifer, D. A. Shell, and M. J. Mataric, "The role of physical embodiment in human-robot interaction," in *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, 2006, pp. 117–122.
- [9] J. M. Kory Westlund *et al.*, "Flat vs. Expressive Storytelling: Young Children's Learning and Retention of a Social Robot's Narrative," 2017.
- [10] "Anthropometry and biomechanics." [Online]. Available: <https://msis.jsc.nasa.gov/sections/section03.htm>. [Accessed: 20-Feb-2017].
- [11] N. Endo and A. Takanishi, "Development of Whole-Body Emotional Expression Humanoid Robot for ADL-Assistive RT Services," *J. Robot. Mechatron.*, vol. 23, no. 6, pp. 969–977, 2011.
- [12] T. Kishi *et al.*, "Development of a Humorous Humanoid Robot Capable of Quick-and-Wide Arm Motion," *IEEE Robot. Autom. Lett.*, vol. PP, no. 99, pp. 1–1, 2016.
- [13] K. R. Scherer, "What are emotions? And how can they be measured?," *Soc. Sci. Inf.*, vol. 44, no. 4, pp. 695–729, Dec. 2005.
- [14] K. R. Scherer, "Toward a dynamic theory of emotion," *Geneva Stud. Emot.*, vol. 1, pp. 1–96, 1987.
- [15] P. Ekman and W. Friesen, *Facial Action Coding System*. Consulting Psychologists Press, Stanford University, Palo Alto, 1977.
- [16] D. Philosophia, "Johannalie Susanna Knoetze," University of Pretoria, 2012.
- [17] P. Ekman, "Are there basic emotions?," *Psychol Rev.*, vol. 99, pp. 550–3, 1992.
- [18] A. Ortony and T. J. Turner, "What's basic about basic emotions?," *Psychol. Rev.*, vol. 97, no. 3, p. 315, 1990.
- [19] R. Plutchik and H. Kellerman, *Emotion: theory, research, and experience. Vol. 1, Theories of emotion*. New York: Academic P., 1980.
- [20] P. Ekman, "Basic Emotions," in *Handbook of Cognition and Emotion*, T. D. R. Scientist and M. J. P. of C. Psychology, Eds. John Wiley & Sons, Ltd, 1999, pp. 45–60.
- [21] B. Weiner and S. Graham, "An attributional approach to emotional development," in *Emotions, cognition, and behavior*, 1984, pp. 167–191.
- [22] W. James, "What is an emotion?," *Mind*, vol. 9, no. 34, pp. 188–205, 1884.
- [23] J. A. Russell, "A circumplex model of affect," *J. Pers. Soc. Psychol.*, vol. 39, no. 6, pp. 1161–1178, 1980.
- [24] P. J. Lang, M. K. Greenwald, M. M. Bradley, and A. O. Hamm, "Looking at pictures: Affective, facial, visceral, and behavioral reactions," *Psychophysiology*, vol. 30, no. 3, pp. 261–273, 1993.
- [25] H. Cai and Y. Lin, "Modeling of operators' emotion and task performance in a virtual driving environment," *Int. J. Hum.-Comput. Stud.*, vol. 69, no. 9, pp. 571–586, Aug. 2011.
- [26] K. Markov and T. Matsui, "Music Genre and Emotion Recognition Using Gaussian Processes," *IEEE Access*, vol. 2, pp. 688–697, 2014.
- [27] C.-H. Hjortsjö, *Man's face and mimic language*. Studen litteratur, 1969.
- [28] "Microsoft Cognitive Services." [Online]. Available: <https://www.microsoft.com/cognitive-services/en-us/>. [Accessed: 22-Feb-2017].
- [29] N. Gourier, D. Hall, and J. L. Crowley, "Estimating face orientation from robust detection of salient facial structures," in *FG Net Workshop on Visual Observation of Deictic Gestures*, 2004, vol. 6.
- [30] D. Lundqvist, A. Flykt, and A. Öhman, "The Karolinska directed emotional faces (KDEF)," *CD ROM Dep. Clin. Neurosci. Psychol. Sect. Karolinska Institutet*, no. 1998, 1998.